Duplicate Record Merge: Patient Merge
Release Notes
Patch XT*7.3*23

Patient Merge provides an automated method to eliminate duplicate patient records within the **V*IST*A** database. It is an operational implementation of the Duplicate Resolution Utilities, which were released to the field with Kernel Toolkit.

The overall process consists of three major subject areas:

1. search for potential duplicate record pairs,
2. review, verification, and approval of those pairs, and
3. the merge process

SEARCH FOR POTENTIAL DUPLICATE RECORD PAIRS

- The search process is available for initiation by IRM personnel who hold the XDRMGR key.
- The search process is lengthy and is recommended for off-peak hours.
- Utilities are available for pausing and restarting the search process.
- The Patient Merge application performs comparisons on key fields in the PATIENT file (#2).
- A Basic search initiates a search on a file for the first time.
- Routines are activated containing preset specifications for duplicate tests.
- These tests are identified in the DUPLICATE RESOLUTION file (#15.1).
- Each test uses its corresponding field and file numbers to assist in identifying potential duplicate records. A value is ascribed to the field based upon similarity in the data present. Those fields are:

Name	Last Separation Date
Social Security	(Last Discharge Date)
Sex	Mother's Maiden Name
Date of Birth	Claim Number
Date of Death	

- All field values are tabulated to form a single value, which is compared against the Threshold Percentage (a site parameter).
- When record pair scores evaluate above the Threshold, they are considered potential duplicates and entered into the DUPLICATE RECORD file (#15) for further processing.

REVIEW, VERIFICATION, AND APPROVAL OF THOSE PAIRS

Review and Verify Potential Duplicates

- Once a potential duplicate pair has been found, the process of reviewing and verifying record pairs begins.
- The review and verification process may begin while the search is running.
- A site may determine reviewers based upon their business practices.
- The review and verification process includes two levels of review:

    1. Primary reviewer– This task was initially seen as an MAS responsibility. The primary reviewer performs a review of patient demographic information to determine if the pair represents a duplicate record. If so, the primary reviewer selects the merge direction.
    2. Ancillary reviewers– If data from ancillary services is present, notification (via MailMan message or alert– or both) may be sent to those designated as ancillary reviewers.

- Reviewers determine whether the record pair is a duplicate, not a duplicate (so that subsequent processing need not occur), or that they are unable to determine the status.
- Where appropriate, reviewers may mark data to be overwritten.
- Those record pairs that are determined to be verified duplicates are marked as such and are then available for approving to be merged.

Approve Verified Duplicates for Merging

The intent of the approval step is to ensure that a conscious decision be made in making verified duplicate record pairs available for a merge process. The following factors apply:

- All verified record pairs, or selected pairs, can be approved.
- The approval step follows a site defined waiting period.
- Reviewers are responsible for approving verified duplicates.

THE MERGE PROCESS

- The merge process is available for initiation by IRM personnel who hold the XDRMGR key.
- All approved record pairs are included in a merge process when scheduled.
- The merge process is lengthy and is recommended for off-peak hours. It can vary in length of time depending upon the number of parallel jobs (threads) selected.
- Utilities are available for pausing and restarting the merge process.
- A routine, known as the validator, examines data in the designated pairs to find potential data errors. If problems are found, these records will not be merged. Notification will be sent to an associated mailgroup.
- The merge process merges verified duplicate records in the following order: first, files that require special handling, then the primary file, then the resolution of pointers.
- The resolution of pointers for the primary file or any of those involving special processing involve three phases, described below:

    * The first two phases permit identification of entries requiring modification based on their IENs (DINUMed) or by cross-references and are fairly rapid.
    * The third phase involves all other pointers and can be lengthy.

- Several special processing routines have been written to handle those database entries that point to the PATIENT file in an unusual manner.
- Entries for each special processing routine have been made in the PACKAGE file multiple, Affects Record Merge (Field #20).
- A stub record is maintained in order to disallow reuse of PATIENT file internal entry numbers.
- The merge process is a background job that involves moving data and repointing information. Be aware that it should not be running when changes are being made to Data Dictionaries or when data conversions are taking place.

"Before Image" Global Designed for Archiving

Concurrent with the merge, entries are made in a new global for each record making up the pair. The entries are intended to provide a "before-merge" image. The establishment of a separate global facilitates the ability to remove it from the active system and save it off as an archive.

Please note that the merge is a non-reversible process. Once the pair of records is merged, there is no automated way of undoing the process.

USE OF SYSTEM RESOURCES

Some processes incorporated in Patient Merge are CPU intensive jobs. The length of time for each process is dependent upon factors such as the number of entries in files and the system configuration. Utilities are specifically provided to allow for pausing and restarting these jobs. These utilities are capable of knowing the last action completed so that an entire job does not need to be reinitiated.

- The search process involves the PATIENT File comparisons and has been known to exceed 100 hours.
- A merge process can range from approximately 15 hours to more than 24 hours.

Multiple Parallel Processing

The Patient Merge application has been written to support multiple parallel jobs (threads - as specified by the site) during the merge process. However, decreased overall processing time is exchanged for increased system utilization.

DUPLICATE RECORD MERGE: PATIENT MERGE

RELEASE NOTES

PATCH XT*7.3*23

April 1998

Department of Veterans Affairs
Software Service
Infrastructure Product Line